评论簇在网络舆论中的情感倾向代表性研究*

杨小平 马奇凤 余 力 莫雨婷 吴佳楠 张 悦
(中国人民大学信息学院 北京 100872)

摘要: 【目的】验证评论簇在网络舆论中具有情感倾向代表性作用。【方法】提出基于社会网络分析的评论簇对象情感倾向性分析模型。针对网络新闻事件, 以新闻的用户评论(评论总集)为语料数据, 对语料数据进行结构化处理和分析, 借助评论主体的形式化关系, 建立具有网络节点和拓扑连接关系的知识图谱, 寻找最优评论簇。以评论簇的评论主体及其对应的评论对象为主, 对评论簇内的核心人物及其评论进行语义分析, 计算得到评论簇情感倾向, 并与对应新闻的评论总集情感倾向作对比。【结果】实验结果表明, 评论簇和评论总集中的情感强度趋于一致, 新闻的评论簇对新闻具有较好的情感倾向代表性, 并能将网络舆情对象情感挖掘算法的性能提高 58%。【局限】由于本文的评论簇对象情感倾向性分析模型在情感特征词识别和抽取方法上使用不够完善, 导致少量中文分词和词性标注错误、语法依存关系错误, 且未将程度词考虑在内。【结论】评论簇在网络舆论中具有情感倾向代表性作用, 可提高网络舆情对象情感计算的性能, 可灵活有效地降低舆情分析的时间和空间复杂度。

关键词: 语义网络 知识图谱 核心人物 网络舆情 评论簇 情感倾向计算
分类号: TP391

1 引 言

网络舆论是公众在特定时间里通过各种网络渠道对新闻事件进行观点表述和思想传播而形成的情绪总和。网络舆论在多数情况下是以多样化的评论总集为发展根源, 一条新闻事件的评论总集呈现出网民对于该事件的态度, 亦称情感倾向。同时, 被大多数网民认可或者反驳的评论则会以点赞或回复形式形成聚集的“评论簇”, “评论簇”就成为舆论情感倾向研究的核心。如何有效把握网络舆论发展的势态, 即网络舆情, 评论簇就成为网络舆情中情感倾向代表的重要组成部分。评论对象是评论主体对某新闻进行评论时产生的评论文本主语。评论簇研究的关键任务是挖掘评论领袖[1]即后文提出的核心人物或核心评论主体、挖掘评论对象以及分析评论情感倾向等。本文针对这一研究点, 利用新闻事件的评论语料, 采用社会网络分析[2]的语义网络知识图谱技术和网络舆情情感挖掘分析算法[3]等相关知识, 提出基于社会网络分析的评论簇对象情感倾向性分析模型。

2 相关研究

网络舆论的情感倾向代表性研究方向主要包括: 以社会网络分析(SNA)为基础而形成的对核心人物挖掘的网络舆情研究; 以机器学习为主的语料文本挖掘分析, 最终形成情感分析算法。

(1) 以 SNA 为分析方法的主要通过对网络中传播信息的节点构成的网络进行研究, 发现该网络的结构特征, 定量计算得到语义关系网络中的核心人物, 助

通讯作者: 余力, ORCID: 0000-0001-8503-2535, E-mail: buaayuli@ruc.edu.cn。

*本文系北京市教育委员会科技计划面上项目“文本挖掘若干关键问题的研究”(项目编号:KM201511232016)和国家自然科学基金项目“推荐–采纳模式下的病毒营销用户影响研究”(项目编号: 71271209)的研究成果之一。

于分析舆情信息传播模式。石彭辉[1]研究SNA在网络舆情方面的应用，主要通过SNA技术中的参数分析，来进行舆情研究。刘继等[2]对网络舆情传播模式中单关键点型、多关键点型、链式型传播模式进行分析，对网络结构中的强势节点和桥节点的作用进行讨论。赵德伟等[4]以某网络热点话题为例，对该话题社会网络进行整体挖掘，计算分析点度中心度、密度、直径、聚类系数等参数信息，并提出舆情监管建议。

(2) 以机器学习为主则主要是对网络传播的信息内容进行语义分析研究，发现重要网络舆情信息。杜嘉忠等[5]提出一种基于领域专用情感词的网络评论情感分析方法。该方法构建特征–情感词本体，利用本体对网络上的产品评论进行情感分析，并与基于Senti-HowNet词典的情感分析方法进行对比。韩瑞凯[6]在微博情感的特征产生与选择以及分类器研究方面，主要介绍了基于朴素贝叶斯微博情感分析情况，针对微博短文本的特点，将微博视为单一观点和观点分割两种情况进行分析。

综上，从社会网络分析和语义情感计算两个分支的相关研究现状可以看出，社会网络分析可以从非机器学习角度对语义情感计算提供相关指标的快速分析，二者之间有很高的关联度；语义情感分析的相关核心指标可以借助社会网络分析方法实现。本文采用社会网络分析和文本挖掘情感倾向研究相结合的方法，通过对网络舆论中的评论簇进行情感倾向分析并建立评论簇和评论总集的关联关系，实现评论簇对网络舆论具有较好的情感倾向代表的可行性和有效性验证。

3 基于社会网络分析的评论簇对象情感倾向性分析模型

本文在首次实验过程中对短文本进行处理分析，由于评论文本属于短文本，情感特征稀疏，且评论对象抽取难度较大，SVM情感分类器[7]对此类评论文本的处理效果不及本文的研究模型效果理想，故在二次实证过程中提出基于社会网络分析的评论簇对象情感倾向性分析模型，该模型采用社会网络分析的相关知识推理算法和网络舆情情感挖掘计算，进行多次重复性实验，验证了评论簇在评论总集中具有显著的情感倾向代表性特征。

社会网络分析是对网络结构及属性特征，包括网络中的个体属性和整体属性进行分析[8]。本文以形成网络舆论的评论总集为基础数据源，运用SNA的定量计算指标挖掘出核心人物，即本文研究过程中的评论主体；以评论主体为主构成的传播信息节点[9]的语义网络知识图谱为基础得到评论主体和评论主体对应的评论簇；再利用网络舆情情感倾向分析的相关算法，包括文本预处理、特征抽取、评论对象的情感结构化表示和情感倾向分析，将评论簇和评论总集中带有情感色彩的主观性评论文本进行语义情感倾向性分析并将两者作对比。在本文研究过程中针对上述模型做了多次重复实验，并对实验中指标参数以及算法进行多次改进得出理想的结果，通过评论簇的情感倾向较准确地得出网络舆情的情感倾向，除了在分析计算性能上提高之外，更能为舆情热度、监控和预测作支撑。

本研究模型的两大关键步骤分别是评论簇定量计算和评论簇情感倾向性计算。

3.1 评论簇定量计算

评论簇定量计算是本文模型的关键步骤所在，是为寻找到评论总集中所有的核心评论主体(核心人物)，进一步从核心评论主体得到与核心评论主体对应的评论簇 C_1。评论簇定量计算是基于新闻的大量非结构性评论文本总集C，采用社会网络分析和知识推理的相关理论，根据评论主体互相之间的关联性构建的知识图谱，利用知识图谱中节点的威望度、中心度、评论质量计算得出处于网络核心和发言较活跃的人物，并根据核心人物得到与核心人物相关的网络关系图，以及图中每个节点对应的评论主体及评论，即可寻找到最优评论簇 C_1；由于新闻评论的“顶”、“踩”数量对评论的观点倾向性影响较大，故抽取“顶”、“踩”数≥N的评论作为评论簇 C_2，评论簇 C_2 的阈值设为N。其中，判断知识图谱中核心评论主体的指标分别为威望度、中心度、评论质量。

(1) 威望度

根据Web原评论页面抽取出的量级元数据，运用SNA方法得到某节点入度与所有网络节点入度的比值后进行排序，得出核心人物的威望度[6]。入度是所有指向该节点的弧个数之和，在本文中指某条评论被回复的个数。运用比值法求得威望度作为核心人物衡量

指标，节点的入度与威望度成正比，一个节点的入度越高，则威望度越高，那么该节点所代表用户的评论内容被其他用户回复的次数就越多，该用户在论坛中所处的位置就越重要。威望度的计算公式[6]如下：

$$P_D(v_i)=\frac{x_{i\text{-}}}{\sum_{i=1}^{k}x_{i\text{-}}} \qquad (1)$$

其中，$x_{i\text{-}}$表示节点 v_i的入度。

(2) 中心度

一个节点的中心度越高，表示该节点所代表的评论主体与其他评论主体的互动就越多，说明该评论主体在论坛中越活跃，活跃的评论主体可以带动整个网络的评论能动性。在此思路的指导下，依据 SNA 技术，借助于 Pajek 分析平台和人工分析计算对比得到标准化处理的网络中心度数据，进一步得到网络中活跃度高的评论主体。

(3) 评论质量

评论质量[10]用 M 表示，由于不同评论主体的回复数差别较大，因此要进行归一化处理，如公式(2)。定义评论质量是为了更加精准地得到核心人物，作为寻找正确核心人物的依据，区别于核心人物威望度，评论质量是相对于局部社会网络中评论主体因受回复数影响而言，其回复数与其他回复数的标准化比较，权重值越大的点则评论质量越高。

$$M=\frac{m-m_{min}}{m_{max}-m_{min}} \qquad (2)$$

其中，m 为某一评论主体所得回复数，m_{max} 为所有评论主体中最大回复数，m_{min} 为所有评论主体中最小回复数。

3.2　评论簇情感倾向性计算

本文模型中评论簇情感倾向性计算阶段主要是把前期评论簇定量计算阶段得到的非结构化评论文本转化为结构化评论文本，将评论文本中的评论对象和情感特征词抽取并进行结构化表示，通过语义分析算法做评论文本情感倾向分析。其中，评论文本包括评论总集和评论簇两部分。此阶段分析步骤主要有：文本预处理、特征抽取、评论对象的情感结构化表示、评论对象的情感倾向分析。

(1) 文本预处理

文本预处理的主要目的是将非结构化评论文本总集和评论簇分别转化为结构化评论数据，有利于后续实验的二元组分析。文本预处理算法的核心步骤是：将每个评论段落分为若干语句，将语句按标识符分割成若干短语词组。包括采用 ICTCLAS 程序进行分词和词性标注[11]、去停用词、词频统计和共现词合并分析，对评论总集和定量计算得出的评论簇分别做文本预处理。

(2) 特征抽取

本文的特征抽取是指新闻评论文本中评论对象的抽取和评论对象对应的情感特征词抽取[12]。特征抽取算法的核心步骤是：根据文本预处理得到的词频统计、共现词合并和语法词性分析[13]，得到新闻的评论对象；再进行句法分析和依存关系分析得出合理的短语搭配模式来匹配每个对象的观点特征词。本文的每一短文本都是以评论对象为标识以抽取能代表评论对象情感的语义词，以<动词/v>、<形容词/a>和<名词/n>为情感倾向特征词[14]，得到具有评论对象、特征观点的二元组结构表[15]：

$$\text{Lexicon}=(\text{Object, Feature}) \qquad (3)$$

其中，Object 指抽取的评论对象，Feature 指能代表评论对象的情感倾向特征词。

(3) 评论对象的情感结构化表示

对上述所得的二元组结构表，使用大连理工大学信息检索研究室的情感词汇本体[16]进行极性和程度匹配，得到四元组结构：

$$\text{Lexicon}=(\text{Object, Feature, Polarity, Degree}) \qquad (4)$$

其中，Polarity 指 Object 对应 Feature 在情感词汇本体的极性，Degree 指 Object 对应 Feature 在情感词汇本体中的极性程度。

(4) 评论对象的情感倾向分析

将正向情感强度值和负向情感强度值计算分别表示，若将二者做合差运算，只能反映出情感对象最终情感倾向[17]的一维特性，不利于直观反映出评论对象的正面情感强度值和负面情感强度值的二维特性，正负值分开计算更有利于实验的具体性和反映评论文本的客观性；由于评论簇 C_1 和 C_2 的重要性也会影响实验的准确率，对评论簇 C_1 和 C_2 的权重值θ在实验过程中进行参数调节，取得合理的θ值以提高实验的准确率。基于上述依据和四元组结构表公式(4)，利用情感计

算公式(5)，得出评论簇中评论对象的情感倾向强度值。

$$SO(Object)=\begin{cases}\sum_{i=1}^{n}(Obj,pword)=n\dfrac{polarity(i)\times degree(i)}{n+m}\theta & \theta=\alpha或\beta\\ \sum_{j=1}^{m}(Obj,nword)=-m\dfrac{polarity(j)\times degree(j)}{n+m}\theta & \theta=\alpha或\beta\end{cases} \tag{5}$$

其中，SO(Sentiment Orientation)表示评论对象对应情感特征词的情感强度值；pword 和 nword 分别指正向情感倾向词和负向情感倾向词；n 和 m 分别指正向情感倾向词和负向情感倾向词的总个数，由于正向情感词和负向情感词所占比例不同，为了真实反映情感倾向值的正确分布，公式(5)中加入正、负情感词在总情感词集中的权重比例；polarity(i)为情感倾向词的极性；degree(i)为情感倾向词的程度；θ为评论簇的权重因子，取值为α或β，当情感倾向词在评论簇 C_2 时取值为α，当情感倾向词在评论簇 C_1 时取值为β。

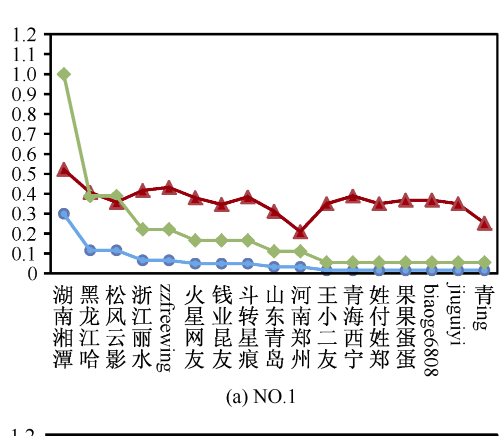
4 实证研究与结果分析

从三个不同领域的热点话题选取具有代表性的新闻评论语料，分别为“成都被打女司机” 2 802 条评论语料、“复旦投毒案” 12 717 条评论语料、“王菲谢霆锋下月大理成婚” 30 959 条评论语料，分析步骤如下：

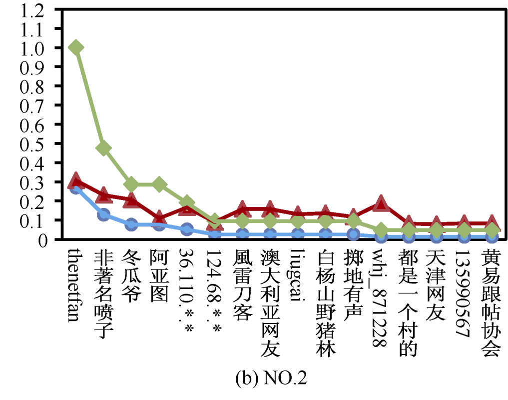
4.1 评论簇抽取

针对每一个特定新闻页面的所有评论数据总集，抽取出新闻跟帖评论中所有评论主体的特征项：评论主体标识符、IP 地址、评论内容以及“顶”、“踩”数，每一个特定新闻页面所有评论总集 C 作为实验分析的对比源数据。每个评论主体对应一个节点，建立具有网络节点和拓扑连接关系的知识图谱，以每个节点的威望度、中心度以及评论质量三个参数寻找核心人物节点。图 1 为三个数据总集节点在分析过程中对应的三个参数的数值分析，由于评论主体数量较多难以全部呈现，图中仅列出部分。

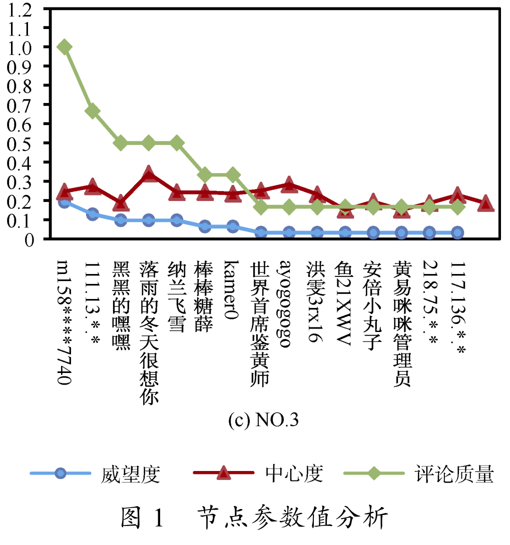
根据上述三个参数分析可得到核心人物节点；根据已选核心人物节点从知识图谱[18]中找到与之相关的所有关系网络，图 2 为具有代表性的网络关系图；结合参数分析和网络关系图找到核心人物节点在评论文本中对应的评论主体及评论，即寻找到最优评论簇 C_1。



(a) NO.1



(b) NO.2



(c) NO.3

图 1 节点参数值分析

对于抽取“顶”、“踩”数为 N 的评论作为评论簇 C_2。由于热点新闻评论贴中“顶”、“踩”数靠前的评论都是按照“顶”数降序排列，本实验以“顶”数作为单一阈值分析指标，且根据大量评论文本的“顶”数据作分析得到阈值 N=200，寻找到最优评论簇 C_2。

由于评论簇 C_1 和 C_2 没有客观固定的实验权重值作依托，针对此问题进行重复性实验以确定此参数因子，依据理论设定 8 组实验参数值，做α, β生成参数调

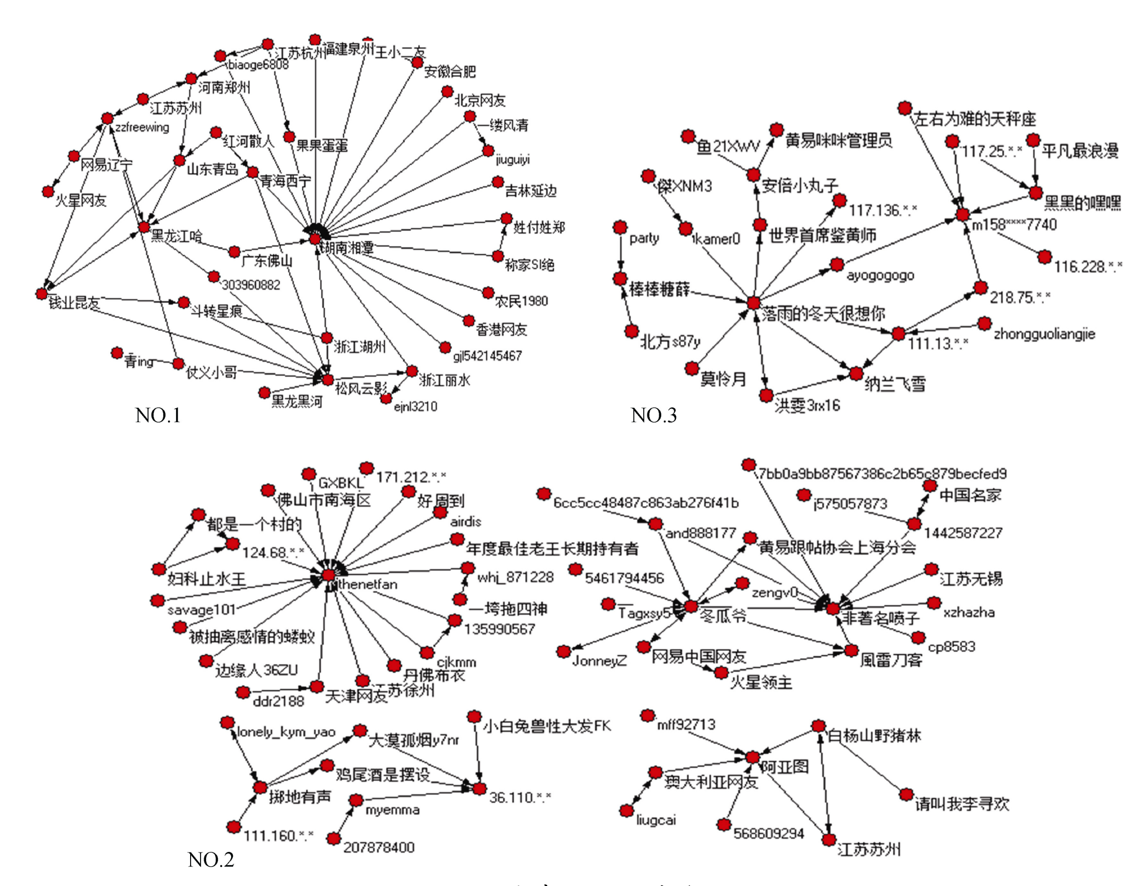


图 2 代表性网络关系图

优实验。根据 8 组参数分别计算得到三个评论簇集的一维对象情感倾向值，并与各自对应的评论总集 C 的一维对象情感倾向值作对比，以欧氏距离作为 C 和 $C_i(i=1,2)$对比的评价函数 d(见公式(6))，对比后分别求得三组数据平均值 $\bar{d}$ 寻求到α，β的最优值(见表 1)。

$$d=\sqrt{(X_2-X_1)^2+(Y_2-Y_1)^2} \quad (6)$$

表 1 α，β生成参数调优实验

实验	α	β	$\bar{d}$
基准实验	1	1	1.052093152
去参实验	0	1	1.679315337
	1	0	1.652452722
参数寻优实验	0.5	0.5	0.832346082
	0.6	0.4	0.990454441
	0.7	0.3	0.845990544
	0.8	0.2	0.697710542
	0.9	0.1	1.350740538

根据理论取得上述 8 组数据作为参数调节的实验数据。本次实验表明，当α=0.8，β =0.2 时，评论簇计算得到的对象情感倾向值和评论总集计算得到的对象情感倾向值最为接近，即将公式(5)中的α确定为 0.8，β确定为 0.2。

4.2 评论簇情感倾向性计算

对所得评论簇 C_i 进行分词、词性标注、去停用词、词频统计、共现词合并分析和语法词性分析得到评论文本中的主要评论对象。评论文本中的评论对象可能会出现“他” “她”类似的代词，根据人民大学信息学院文本挖掘与数据分析课题组自行开发的指代消解工具对指代词进行一致化处理后再进行句法分析和依存关系分析，得到主要评论对象对应的以<动词/v>、<形容词/a>和<名词/n>为主要情感倾向词的二元组关系表，再使用大连理工大学信息检索研究室的情感词汇本体进行极性和程度匹配，得到四元关系表，部分代表性特征项属性关系见表 2。与此同时，分别对三则新闻的评论总集源数据 C 进行上述相同步骤分析。在识别评价对象对应的情感特征词的个别错例现象中发现，其主要错误原因是程序优化不够使得短文本没有得到正

确的语法依存关系分析，在遇到多个评价的复杂句中评价对象与情感特征词距离过远或不同评价对象之间的特征词配对错误，会造成个别二元组的配对误差。此原因造成误差的数据在进行二元组存储时根据本课题组组内研究已有成果中的SentiRuc情感词典进行一次性修正，少数情感特征词在本体词库匹配失败，则利用 HowNet 进行同义词替换，尽可能取得最小误差的目标数值。

表 2　对象–情感特征词四维关系表

组别	Object	Feature	Polarity	Degree	Object	Feature	Polarity	Degree
NO.1	女司机	慈善	1	5	女司机	作死	2	9
	女司机	健康	1	3	女司机	违规	2	5
	男司机	顶天立地	1	7	女司机	可怜虫	2	3
	男司机	可耻	2	5	男司机	凶暴	2	7
NO.2	林森浩	杀人灭口	2	7	法律	沦丧	2	5
	林森浩	亏欠	2	7	法律	公道	1	5
	林森浩	毒害	2	7	黄洋	含冤	2	5
	林森浩	自私	2	7	黄洋	无辜	1	1
NO.3	王菲	任性	2	3	谢霆锋	逍遥	2	5
	王菲	辜负	2	3	谢霆锋	不闻不问	2	1
	王菲	潇洒	1	5	张柏芝	无私	1	5
	王菲	祝福	1	9	张柏芝	矫情	2	3

4.3　实验结果对比分析

基于上述表示的四元组结构表，利用情感计算公式(5)，分别计算出评论总集 C 和评论簇 C_i 的评论对象在各自对应新闻网络舆论中的情感强度值，再将上述计算出来的各自对应的评论总集 C 和评论簇 C_i 的情感强度值进行对比，结果如图 3 所示：

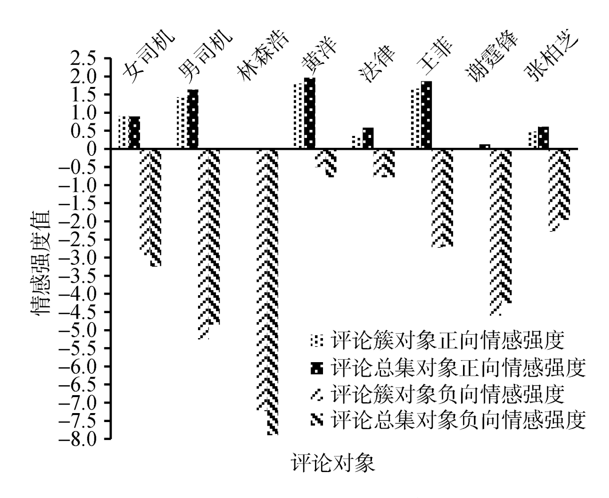


图 3　评论总集和评论簇的情感强度值

图 3 表明，三个新闻评论文本的源数据集通过分析处理，利用本文模型计算出各自对象的情感强度值，将评论簇 C_i 和评论总集 C 各自评论对象的情感倾向强度值作对比。新闻 1 中评论文本的主要评论对象为“女司机”、“男司机”；新闻 2 中评论文本的主要评论对象为“林森浩”、“黄洋”、“法律”；　新闻 3 中评论文本的主要评论对象为“王菲”、“谢霆锋”、“张柏芝”，三则新闻采用社会网络分析的评论簇对象情感倾向性分析模型的计算结果，可以得出以下结论：

(1) 通过反复改进本模型实验过程中的不合理和疏漏问题，所选取的不同领域热点话题中具有代表性的三则新闻的评论簇情感强度值和文本总集情感强度值接近度保持在理想误差范围内。在同样的源数据集下，SVM 情感分类器对此类评论文本的处理效果不及本文的模型效果理想。从实证角度看，本研究的论据、研究模型是可行的，且效果理想。

(2) 目前，网络舆论中评论对象情感倾向的通用计算是根据评论总集的源数据计算而得到。通过图 3 可以看出，三则新闻网络舆论中，各自评论对象的正负情感强度在评论簇和评论总集两者中趋于一致，如果直接采用评论簇分析网络舆情的情感倾向强度值，则在计算机处理耗能方面，性能可提高 58%，可以有效降低网络舆情分析的时间和空间复杂度。

(3) 根据理论，代表性强的舆论反映了整体评议的共同取向[19]。此次实验验证了评论簇在网络舆论中

具有理想的情感倾向代表性作用，进而对深入研究网络舆情的共同取向分析具有一定的理论和实践意义。

5 结 语

本文基于大量评论语料，提出基于社会网络分析的评论簇对象情感倾向性分析模型，采用SNA和网络舆情情感挖掘算法建立了一种评论对象及其对应评论特征的抽取规则，根据这种规则较好地计算出评论对象的情感倾向强度，验证了评论簇在网络舆论中的代表性理论。本文基于真实的新闻评论文本源，运用了正确的理论和方法，以保证整个研究的真实性和可靠性。因此，在评论文本分析中直接采用评论簇代表评论总集，在计算机处理耗时方面，性能可提高 58%，可有效提高网络舆情分析的效率，对深入研究网络舆情具有有效的理论指导和实践意义。同时，由于本文的评论簇对象情感倾向性分析模型在情感特征词识别抽取方法方面使用不够完善，导致少量中文分词和词性标注错误、语法依存关系错误，且未将程度词考虑在内。在后续研究中，除了优化该算法，还考虑将评论文本中修饰评论对象的程度词列入情感强度计算范围之内，并进一步优化更深层次的中文分词和词性标注，丰富完善本体词库的词语量，使得后续研究体系更加完整，提高情感信息抽取的完整性和情感强度值的精确性。

参考文献：

[1] 石彭辉. 基于社会网络分析的网络舆情实证研究[J]. 现代情报, 2013, 33(2): 27-31. (Shi Penghui. Empirical Studies of Network Public Opinion Based on Social Network Analysis [J]. Journal of Modern Information, 2013, 33(2): 27-31.)

[2] 刘继, 李磊. 基于微博用户转发行为的舆情信息传播模式分析[J]. 情报杂志, 2013, 32(7): 74-77. (Liu Ji, Li Lei. Analysis of Public Opinion Propagation Mode Based on Repost Behavior of Microblog Users [J]. Journal of Intelligence, 2013, 32(7): 74-77.)

[3] 林江豪. 中文微博情感分析关键技术研究[D]. 广州: 广东外语外贸大学, 2013. (Lin Jianghao. Research on Key Techniques of Chinese Micro-blog Sentiment Analysis [D]. Guangzhou: Guangdong University of Foreign Studies, 2013.)

[4] 赵德伟, 徐正巧.基于社会网络分析的网络舆情数据挖掘[J]. 福建电脑, 2014, 15(8): 15-16, 50. (Zhao Dewei, Xu Zhengqiao. Network Public Opinion Data Mining Based on Social Network Analysis [J]. FuJian Computer, 2014, 15(8): 15-16, 50.)

[5] 杜嘉忠, 徐健, 刘颖. 网络商品评论的特征–情感词本体构建与情感分析方法研究[J]. 现代图书情报技术, 2014(5): 74-82. (Du Jiazhong, Xu Jian, Liu Ying. Research on Construction of Feature-Sentiment Ontology and Sentiment Analysis [J]. New Technology of Library and Information Service, 2014(5): 74-82.)

[6] 韩瑞凯. 基于社区发现的网络舆论导向系统研究与应用[D]. 北京: 北京交通大学, 2010. (Han Ruikai. Research and Application of Network Consensus Guidance System Using Community Detection [D]. Beijing: Beijing Jiaotong University, 2010.)

[7] 肖正, 刘辉, 李兵. 一种基于语义距离的 Web 评论 SVM 情感分类方法[J]. 计算机科学, 2014, 41(9): 248-252, 284. (Xiao Zheng, Liu Hui, Li Bing. SVM Sentiment Classifier Based on Semantic Distance for Web Comments [J]. Computer Science, 2014, 41(9): 248-252, 284.)

[8] 黄晓斌, 赵超. 文本挖掘在网络舆情信息分析中的应用[J]. 情报科学, 2009, 27(1): 94-99. (Huang Xiaobin, Zhao Chao. Application of Text Mining Technology in Analysis of Net-Mediated Public Sentiment [J]. Information Science, 2009, 27(1): 94-99.)

[9] 李卓卓, 丁子涵. 基于社会网络分析的网络舆论领袖发掘——以大学生就业舆情为例[J]. 情报杂志, 2011, 30(11): 67-70. (Li Zhuozhuo, Ding Zihan. Exploring Online Opinion Leadership Based on Social Network Analysis-Public Opinion of College Student Employment Taken for Example [J]. Journal of Intelligence, 2011, 30(11): 67-70.)

[10] 聂卉. 基于内容分析的用户评论质量的评价与预测[J]. 图书情报工作, 2014, 58(13): 83-89. (Nie Hui. Content-oriented Evaluation and Detection for Product Reviews[J]. Library and Information Service, 2014, 58(13): 83-89.)

[11] ICTCLAS 汉语分词系统[DB/OL]. [2013-07-02]. http://ictclas.nlpir.org. (ICTCLAS Chinese Segmentation System [DB/OL]. [2013-07-02]. http://ictclas.nlpir.org.)

[12] 刘建毅, 王菁华, 王枞. 基于语言网络的关键词抽取[C]. 见: 第三届全国信息检索与内容安全学术会议, 2008. (Liu Jianyi, Wang Jinghua, Wang Cong. Keyword Extraction Using Language Network [C]. In: Proceedings of the 3rd National Conference on Information Retrieval and Information Content Security. 2008.)

[13] 杨经, 林世平. 基于 SVM 的文本词句情感分析[J]. 计算机应用与软件, 2011, 28(9): 225-228. (Yang Jing, Lin Shiping. Emotion Analysis on Text Words and Sentences Based on

SVM [J]. Computer Applications and Software, 2011, 28(9): 225-228.)

[14] 张圣声，阳爱民，周咏梅，等. 微博产品评论的情感倾向性分析方法[J]. 山西大学学报：自然科学版，2015，38(2): 215-222. (Zhang Shengsheng, Yang Aimin, Zhou Yongmei, et al. Method of Sentiment Orientation Analysis for Micro-blogging Product Reviews [J]. Journal of Shanxi University: Natural Science Edition, 2015, 38(2): 215-222.)

[15] 杜伟夫. 文本倾向性分析中的情感词典构建技术研究[D]. 哈尔滨：哈尔滨工业大学，2010. (Du Weifu. Research on Sentimental Lexicon Construction for Text Sentiment Analysis [D]. Harbin: Harbin Institute of Technology, 2010.)

[16] 徐琳宏，林鸿飞，潘宇，等. 情感词汇本体的构造[J]. 情报学报，2008, 27(2): 180-185. (Xu Linhong, Lin Hongfei, Pan Yu, et al. Constructing the Affective Lexicon Ontology [J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 180-185.)

[17] 党蕾，张蕾. 一种基于知网的中文句子情感倾向判别方法[J]. 计算机应用研究，2010，27(4): 1370-1372. (Dang Lei, Zhang Lei. Method of Discriminant for Chinese Sentence Sentiment Orientation Based on HowNet [J]. Application Research of Computers, 2010, 27(4): 1370-1372.)

[18] 杨静，辛宇，谢志强. 基于话题综合因子分析的语义社会网络社区发现算法[J]. 计算机研究与发展，2014，51(3): 559-569. (Yang Jing, Xin Yu, Xie Zhiqiang. Semantics Social Network Community Detection Algorithm Based on Topic Comprehensive Factor Analysis [J]. Journal of Computer Research and Development, 2014, 51(3): 559-569.)

[19] 任韧，纪晓伟，杨斌. 试论网络舆论的特征、作用与掌控方法[J]. 法制与社会，2013 (11): 173-175. (Ren Ren, Ji Xiao Wei, Yang Bin.The Characteristics, Function and Control Method of Network Public Opinion [J]. Legal System and Society, 2013(11): 173-175.)

作者贡献声明:

杨小平：提出研究命题；
余力：设计研究方案；
莫雨婷，吴佳楠，张悦：采集、清洗和分析数据；
马奇凤：实现研究方案，进行实验，论文起草及最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据[1-3]见期刊网络版 http: //www.infotech.ac.cn；支撑数据[4]可通过电子邮件向大连理工大学信息检索研究室索取，E-mail: irlab@dlut.edu.cn。

[1] 杨小平，马奇凤，余力，莫雨婷，吴佳楠，张悦. news_url.rar. 三则不同领域新闻的网页链接.

[2] 杨小平，马奇凤，余力，莫雨婷，吴佳楠，张悦. news-cluster.rar. 新闻评论测试数据集及处理.

[3] 杨小平，马奇凤，余力，莫雨婷，吴佳楠，张悦. paper.rar. 文章中图表数据.

[4] 大连理工大学情感词汇本体库.

收稿日期: 2016-01-25
收修改稿日期: 2016-04-01

Gauging Public Opinion with Comment-Clusters

Yang Xiaoping　Ma Qifeng　Yu Li　Mo Yuting　Wu Jia'nan　Zhang Yue
(School of Information, Renmin University of China, Beijing 100872, China)

Abstract: [**Objective**] The paper examines the role of comment-clusters in public opinion mining. [**Methods**] We proposed a model to study the Comment-Clusters based on social network analysis techniques. First, we collected comments received by online news reports on three trending events as raw data. Second, we analysed structures and contents of these comments with the help of the vector relationship among them to identify the best comment-clusters. Finally, we conducted semantic analysis of the key players and their comments to investigate their sentiments and then compared them with those of the whole data set. [**Results**] The sentiments got from the whole data set and the comment-clusters were very close to each other. Comment-Clusters improved the performance of public opinion mining algorithm. [**Limitations**] The method of identifying and extracting sentiment words might yield errors. **[Conclusions]** The comment-clusters improve the sentiment orientation computing, which helps us obtain the public opinion more efficiently.

Keywords: Semantic network　Knowledge mapping　Key person　Web public opinion　Comment-Clusters　Sentiment orientation computing

---

Tails 和 Innovative 合作为学术图书馆提供一流的解决方案

Innovative Interfaces 已与 Tails 达成合作协议，改善学术图书馆学生的体验，并提供卓越的教学参考书和阅读列表管理工作流。

这一伙伴关系将为 Innovative 和 Tails Aspire 解决方案提供进一步整合的机会，通过图书馆及其数字馆藏，无缝连接学生和教师，支撑他们获取更好的学术成就。随着学术图书馆需求的变化，以及学术环境下图书馆新机遇的出现，这一合作将提升双方公司产品为图书馆带来的价值。

"当我们考虑到技术如何改变人们的学习、资源查找和任务执行的方式时，我们相信学术图书馆未来在高等教育中将比以往发挥更为积极的作用。"Innovative 执行副总裁 Leif Pedersen 说："与 Talis 合作将帮助学术图书馆从自己的馆藏中挖掘出更多的价值，实现更高效率的采购，确保我们的图书馆用户拥有最高品质的解决方案，从而有效服务于他们的机构。"

认识到学术图书馆在高等教育中发挥的宝贵作用，两家公司都渴望继续提供优质的服务，帮助图书馆应对在服务校园师生中所面临的挑战。加强 Talis Aspire 阅读列表和 Sierra 的整合将为双方带来更大的价值，提升图书馆员工工作效率，帮助学生快速方便地获取所需的课程资料。

"我们非常高兴地欢迎 Innovative 加入我们的成长伙伴计划。这是我们使用基于标准的方法来无缝集成我们的高品质服务和重要内容提供商的又一案例。"Talis 商业开发负责人 Mark Bush 说，"我们期待着与 Innovative 共同为我们的欧洲和亚太地区客户谋利。"

(编译自: https://www.iii.com/news-events/pr/talis-and-innovative-partner-deliver-best-class-solutions-academic-libraries)

(本刊讯)